

1. An apparatus for modeling a queue having highly variable arrival rates:  
an input module configured to receive an index of burstiness, a service time, a service coefficient of variation, a server utilization, and a number of servers;  
an arrival module configured to determine an inter-arrival coefficient of variation (COV) using the index of burstiness;  
a COV module configured to determine a COV queue delay comprising the inter-arrival coefficient of variation;  
an exponential (EXP) module configured to determine an EXP queue delay using an exponential distribution of queue arrivals; and  
a weighting module configured to determine a queue delay using the COV queue delay and the EXP queue delay.
2. The apparatus of Claim 1, wherein the weighting module is further configured to determine the queue delay by summing the COV queue delay multiplied by a first factor and the EXP queue delay multiplied by a second factor disproportionate to the first factor.
3. The apparatus of Claim 2, wherein the first factor comprises the server utilization squared and the second factor comprises the amount that unity exceeds the server utilization squared.

4. The apparatus of Claim 2, wherein the first factor comprises the server utilization raised to the power of n, and the second factor comprises one minus the server utilization raised to the power of n, n being a whole number greater than two.

5. The apparatus of Claim 1, wherein the index of burstiness comprises a value representative of estimated arrival patterns.

6. The apparatus of Claim 1, wherein the arrival module is further configured to determine the inter-arrival coefficient of variation as:

$$CV_{arr} = \text{absolute value } (IB - 1) / 2$$

where:  $CV_{arr}$  = inter-arrival coefficient of variation; and

IB = index of burstiness.

7. The apparatus of Claim 1, wherein the COV module is configured to determine the COV queue delay as:

$$QCV_{delay} = S * (CV_{arr}^2 + U^2 * CV_{ser}^2) / (2 * N * U * (1 - U))$$

where:  $QCV_{delay}$  = COV queue delay;

S = service time;

$CV_{arr}$  = inter-arrival coefficient of variation;

U = server utilization;

$CV_{ser}$  = service coefficient of variation; and

N = number of servers.

8. The apparatus of Claim 1, wherein the EXP module is configured to determine the EXP queue delay as:

$$QEXP_{\text{delay}} = (U^N) * (1 - (U^N))$$

where:  $QEXP_{\text{delay}}$  = EXP queue delay;

$U$  = server utilization; and

$N$  = number of servers.

9. The apparatus of Claim 1, wherein the weighting module is configured to determine the queue delay as:

$$Q_{\text{delay}} = U^n * QCV_{\text{delay}} + (1 - U^n) * QEXP_{\text{delay}}$$

where:  $Q_{\text{delay}}$  = queue delay;

$U$  = server utilization;

$QCV_{\text{delay}}$  = COV queue delay;

$QEXP_{\text{delay}}$  = EXP queue delay and

$n$  = a whole number greater than one.

10. A method for modeling performance of a queue having highly variable arrival rates:
- receiving an index of burstiness, a service time, a service coefficient of variation, a server utilization, and a number of servers;
  - determining an inter-arrival coefficient of variation (COV) using the index of burstiness;
  - determining a COV queue delay comprising the inter-arrival coefficient of variation;
  - determining an exponential queue delay (EXP queue delay) using an exponential distribution of queue arrivals; and
  - determining a queue delay using the COV queue delay and the EXP queue delay.
11. The method of Claim 10, wherein determining a queue delay comprises summing the COV queue delay multiplied by a first factor and the EXP queue delay multiplied by a second factor disproportionate to the first factor.
12. The method of Claim 11, wherein the first factor comprises the server utilization squared and the second factor comprises the amount that unity exceeds the server utilization squared.
13. The method of Claim 11, wherein the first factor comprises the server utilization raised to the power of  $n$ , and the second factor comprises one minus the server utilization raised to the power of  $n$ ,  $n$  being a whole number greater than two.

14. The method of Claim 10, wherein the index of burstiness comprises a value representative of estimated arrival patterns.

15. The method of Claim 10, wherein the inter-arrival coefficient of variation is determined as:

$$CV_{arr} = \text{absolute value } (IB - 1) / 2$$

where:  $CV_{arr}$  = inter-arrival coefficient of variation; and

$IB$  = index of burstiness.

16. The method of Claim 10, wherein the COV queue delay is determined as:

$$QCV_{delay} = S * (CV_{arr}^2 + U^2 * CV_{ser}^2) / (2 * N * U * (1 - U))$$

where:  $QCV_{delay}$  = COV queue delay;

$S$  = service time;

$CV_{arr}$  = inter-arrival coefficient of variation;

$U$  = server utilization;

$CV_{ser}$  = service coefficient of variation; and

$N$  = number of servers.

17. The method of Claim 10, wherein the EXP queue delay is determined as:

$$QEXP_{\text{delay}} = (U^N) * (1 - (U^N))$$

where:  $QEXP_{\text{delay}}$  = EXP queue delay;

$U$  = server utilization; and

$N$  = number of servers.

18. The method of Claim 10, wherein the queue delay is determined as:

$$Q_{\text{delay}} = U^n * QCV_{\text{delay}} + (1 - U^n) * QEXP_{\text{delay}}$$

where:  $Q_{\text{delay}}$  = queue delay;

$U$  = server utilization;

$QCV_{\text{delay}}$  = COV queue delay;

$QEXP_{\text{delay}}$  = EXP queue delay; and

$n$  = a whole number greater than one.

19. An article of manufacture comprising a program storage medium readable by a processor and embodying one or more instructions executable by a processor to perform a method for estimating the queue delay time of queuing networks with highly variable traffic arrival rates, the method comprising:

receiving an index of burstiness, a service time, a service coefficient of

variation, a server utilization, and a number of servers;

determining an inter-arrival coefficient of variation (COV) using the index of burstiness;

determining a COV queue delay comprising the inter-arrival coefficient of variation;

determining an exponential queue delay (EXP queue delay) using an exponential distribution of queue arrivals; and

determining a queue delay using the COV queue delay and the EXP queue delay.

20. The article of manufacture of Claim 19, further comprising determining the queue delay by summing the COV queue delay multiplied by a first factor and the EXP queue delay multiplied by a second factor disproportionate to the first factor.

21. The article of manufacture of Claim 20, wherein the first factor comprises the server utilization squared and the second factor comprises the amount that unity exceeds the server utilization squared.

22. The article of manufacture of Claim 20, wherein the first factor comprises the server utilization raised to the power of n, and the second factor comprises one minus the server utilization raised to the power of n, n being a whole number greater than two.

23. The article of manufacture of Claim 19, wherein the index of burstiness comprises a value representative of estimated arrival patterns.

24. The article of manufacture of Claim 19, further comprising determining the inter-arrival coefficient of variation as:

$$CV_{arr} = \text{absolute value } (IB - 1) / 2$$

where:  $CV_{arr}$  = inter-arrival coefficient of variation; and

IB = index of burstiness.



25. The article of manufacture of Claim 19, further comprising determining the COV queue delay as:

$$QCV_{\text{delay}} = S * (CV_{\text{arr}}^2 + U^2 * CV_{\text{ser}}^2) / (2 * N * U * (1 - U))$$

where:  $QCV_{\text{delay}}$  = COV queue delay;

$S$  = service time;

$CV_{\text{arr}}$  = inter-arrival coefficient of variation;

$U$  = server utilization;

$CV_{\text{ser}}$  = service coefficient of variation; and

$N$  = number of servers.

26. The article of manufacture of Claim 19, further comprising determining the EXP queue delay as:

$$QEXP_{\text{delay}} = (U^N) * (1 - (U^N))$$

where:  $QEXP_{\text{delay}}$  = EXP queue delay;

$U$  = server utilization; and

$N$  = number of servers.

27. The article of manufacture of Claim 19, further comprising determining the queue delay as:

$$Q_{\text{delay}} = U^n * QCV_{\text{delay}} + (1 - U^n) * QEXP_{\text{delay}}$$

where:  $Q_{\text{delay}}$  = queue delay;

$U$  = server utilization;

$QCV_{\text{delay}}$  = COV queue delay;

$QEXP_{\text{delay}}$  = EXP queue delay; and

$n$  = a whole number greater than one.